

Wie Texte standardisiert und sprachtechnologisch qualitätsgesichert werden können

Priv.-Doz. Dr. Melanie Siegel, acrolinx GmbH, Berlin,
melanie.siegel@acrolinx.de

Dipl.-Phys. Roland Schmeling, Schmeling + Consultants GmbH, Heidelberg,
r.schmeling@schmeling-consultants.de

Einführung

Dass die Standardisierung der bedeutendste Ansatz zu gleichzeitiger Qualitätssteigerung und Kostensenkung ist, gilt als hinreichend belegt. Unbekannt sind dagegen die weitgehenden Möglichkeiten der methodischen Textproduktion und die dadurch ermöglichte automatische Prüfung.

Methodische Textproduktion

Standardisierung in der Technischen Dokumentation bedeutet, dass es Regeln gibt, nach denen die Texte produziert werden. Auf der Formulierungsebene müssen dazu präzise und nachvollziehbare Regeln entwickelt werden, die in der Redaktion konsequent angewendet werden, um in Qualität, Übersetzung und Wiederverwendung zu profitieren.

Inhalt, Zielgruppe oder Funktion

Der Inhalt der Information (ihr propositionaler Gehalt) liefert nur ungenügend Anhaltspunkte zur Formulierung; ein Wartungsprozess oder eine Funktionsbeschreibung bedingt auf Satzebene noch keine konkrete Formulierung. Auch die Zielgruppeneigenschaften liefert lediglich ein Spektrum, in dem sich mögliche Formulierungsmuster aufhalten dürfen.

Anders ist es mit der Funktion eines jeden Satzes, seiner „Illokution“: Die Entscheidung für eine Zielangabe, eine Voraussetzung, Folgen bei Nichtbeachtung einer Warnung, für eine bedingte Handlungsaufforderung, ein Resultat oder eine Übersicht über Baugruppen liefert einen hinreichenden Rahmen für die Formulierung. So wird ein Resultat beispielsweise im Perfekt formuliert, eine Handlungsaufforderung im Imperativ. Für die automatische Qualitätssicherung liefert dieser Ansatz präzise Regeln, die computerlinguistisch gut abgebildet werden können.

Die Funktion eines Satzes als Basis für Regeln

Die führende Methode in der Technischen Dokumentation, das Funktionsdesign® von Prof. Schäfflein-Armbruster und Prof. Muthig, geht genau in dieser Weise vor (vgl. auch Ried, Strukturierungsmethoden für die TD, tekomp Jahrestagung 2006). Im Mittelpunkt stehen die so genannten Funktionalen Einheiten, mit denen die Formulierungen verbunden sind und deren optimale Abfolge über Sequenzmuster geregelt wird. So enthält beispielsweise das Sequenzmuster Warnhinweis die Funktionalen Einheiten „Art und Quelle der Gefahr“, „Folgen bei Nichtbeachtung“ und „Maßnahmen zur Gefahrvermeidung“ in genau dieser Reihenfolge (Abb. 1). Für jede der drei Funktionalen Einheiten gelten spezifische Formulierings- und Interpunktionsregeln.


 WARNING	Funktionale Einheit	Formulierung
Offen liegende elektrische Komponenten Lebensgefahr durch Stromschlag! ➤ Schütz K1 abschalten und gegen Wiedereinschalten sichern. ➤ Zweite Person mit dem Sichern des Arbeitsbereiches beauftragen.	← Art und Quelle der Gefahr ← Folgen bei Nichtbeachtung ← Maßnahme zur Gefahrvermeidung ← Maßnahme zur Gefahrvermeidung	Nominalphrase Nominalphrase Imperativsatz mit Infinitiv Imperativsatz mit Infinitiv

Abb. 1: Funktionale Einheiten mit unterschiedlichen Formulierungsmustern und Zeichensetzungen

Kontextsensitive Prüfung

Die Vorteile dieser Methode kommen in vollem Umfang erst zum Tragen, wenn die Redaktionsumgebung nach einem Funktionsdesign modelliert wird. Das heißt: Die Elemente der DTD oder die Formatvorlagen im DTP-Programm sind funktional definiert. Beispiel: "result" und "action" anstatt "para" oder "standard". Die automatische Qualitätssicherung kann dann die hinterlegten Regeln den XML-Elementen bzw. Formatvorlagen zuordnen und den jeweils gültigen Regelsatz verwenden.

Warum gibt es keinen allgemein gültigen Regelsatz?

In über 15 Jahren Erfahrung mit Funktionsdesign haben sich zahlreiche Regeln als generell sinnvoll herausgestellt. Dennoch sind die unternehmensspezifischen und konkreten Regeln eines ausgearbeiteten Funktionsdesigns unterschiedlich, da ein Funktionsdesign-Entwickler nutzungorientierte Aspekte wie Zielgruppenerwartungen, Branche, Corporate Language, aber auch produktionsorientierte Aspekte wie die Struktur der Redaktion berücksichtigen muss.

Beispiele: In der einen Redaktionsumgebung werden bedingte Handlungsaufforderungen benötigt und modelliert, in einer anderen ist dies nicht erforderlich. In der einen Redaktionsumgebung sollen Handlungsaufforderungen mit einer persönlichen Anrede formuliert werden, in einer anderen vor allem kurz mit Infinitiven, in einer dritten wird zwischen Warnhinweisen und sonstigen Handlungsaufforderungen differenziert.

Andererseits haben sich für bestimmte Branchen sinnvolle Regeln herauskristallisiert. Redaktionsleitfäden von Unternehmen derselben Branche sind daher gefragte Dokumente, vor allem, wenn sie nach derselben Methode aufgebaut und daher gut vergleichbar sind. Darüber hinaus ist ein Redaktionsleitfaden nach Funktionsdesign eine wesentliche Grundlage für die automatische Prüfung.

Automatische Prüfung technischer Dokumentation

Die Qualitätssicherung durch automatische Prüfung bezieht sich auf die Faktoren Verständlichkeit, Lesbarkeit, Übersetzbarkeit und Konsistenz. Diese Faktoren stehen in direktem Zusammenhang mit der Qualität von Rechtschreibung, Grammatik, Stil und Terminologie, auf die sich automatische Prüfungen beziehen.

Für die Realisierung der automatischen Prüfung technischer Dokumentation stehen zwei Herangehensweisen zur Verfügung. Bei der ersten wird der erlaubte Sprachumfang definiert, so dass alles, was nicht zum definierten Lexikon und den definierten Strukturen gehört, markiert und dem Benutzer zur Änderung vorgelegt wird. In der anderen Herangehensweise werden typische Fehler modelliert. Während die 1. Herangehensweise strikter ist und dem technischen Redakteur keine Freiheiten lässt, ist die 2. Herangehensweise flexibler und

anpassbarer an die Bedürfnisse des Kunden. Sie geht davon aus, dass nicht jede Firma dieselben Stilrichtlinien hat. Stattdessen müssen die Stilregeln an den Bedarf angepasst werden. Auch firmenintern können unterschiedliche Stilrichtlinien gelten, z. B. für technische Dokumente für die Werkstatt und den Endkunden.

Diese Flexibilität kann auch für die Rechtschreib-Prüfung realisiert werden, indem sie eng mit der Firmen-Terminologie verzahnt wird. Eine Benennung der Firmen-Terminologie darf nicht als Rechtschreibfehler angesehen werden, sondern muss in die Terminologieprüfung eingehen. Dem System unbekannte Wörter können zur Aufnahme in die Terminologie vorgeschlagen werden.

Stilregeln beziehen sich – wo vorhanden – auf die Funktionsdesign-Elemente im Text. So gelten z. B. unterschiedliche Werte für die maximale Satzlänge für den Grundtext, Handlungsanweisungen und Überschriften.



Abb. 2: Rückmeldung des Prüfwerkzeugs acrocheck auf einen Fehler in einem nach Funktionsdesign entwickelten Text

Das Definieren automatischer Prüfregeln basiert auf der Analyse großer Datenmengen. Immer wieder stellt sich heraus, dass Fehler, die in technischen Dokumenten gemacht werden, nicht immer dieselben sind wie die, die in theoretischen Überlegungen entstehen. Daher ist die Basis der Arbeit an Prüfregeln eine Datenanalyse, die mit computerlinguistischen Methoden zur Erfassung der Daten unterstützt wird. Zu diesen Methoden gehört die automatische Suche nach ähnlichen Sätzen, also Sätzen, die sich nur leicht voneinander unterscheiden in den Textdaten (das sogenannte Sentence-Clustering).

Zusammenfassung und Ausblick

Der Nutzen der automatischen Prüfung für den Technischen Redakteur ist vielfältig: Der Prozess des Korrekturlesens wird nachhaltig unterstützt. Die Qualität der Dokumente steigt und damit auch die Kosten für die Übersetzung und der Nutzen für die Anwender. Der Nutzen der automatischen Prüfung hängt einerseits von der Präzision der Regeln ab (ein Anwender darf nicht mit Fehlalarmen belästigt werden), andererseits von der Flexibilität und Anpassbarkeit an Stilrichtlinien, Strukturinformation und Terminologie.

Die auf der Funktion der Information basierende Methode Funktionsdesign liefert präzise Regeln und zusätzlich die Basis für eine Modellierung der Struktur in XML, die eine kontextsensitive Prüfung ermöglicht.

Semantische, also die Bedeutung der Information betreffende Regeln können derzeit nur sehr begrenzt toolgestützt geprüft werden. Hier ist nach wie vor die reflektierte Aufmerksamkeit des Redakteurs gefragt und gleichzeitig ein großes Potenzial für die zukünftige Entwicklung.